Consistency analysis of multidimensional gonio–spectrophotometric measurements in interlaboratory comparisons

A. Ferrero[1], J. Campos[1], B. Bernad[1], A. Pons[1], M. L. Hernanz[1], F. M. Martínez-Verdú[2] and A. Höpe[3,†]

[1]Instituto de Óptica "Daza de Valdés" (IO-CSIC), Agencia Estatal CSIC, c/ Serrano 121, 28006 Madrid, Spain
[2]Department of Optics, Pharmacology and Anatomy, University of Alicante, Carretera de San Vicente del Raspeig s/n 03690, Alicante, Spain
[3]Physikalisch-Technische Bundesanstalt, Bundesallee 100, 38116 Braunschweig, Germany

E-mail: Corresponding author: alejandro.ferrero@csic.es

Abstract. The spectral Bidirectional Reflectance Distribution Function (BRDF) is the key quantity to specify the spectral reflectance of materials for any condition of irradiation and detection, and its characterization is quite important for surfaces with a high dependence on these conditions, such as iridescent coatings. In order to evaluate the calibration and measurement capabilities (CMC) of National Metrology Institutes (NMI) with the ability to measure the spectral BRDF, a case study interlaboratory comparison is in progress. Spectral BRDF has both spectral and geometric dependence, and this multidimensionality must be treated in the comparison to provide useful information to the participants about their CMCs. A data analysis method for the comparison is presented in this work, which was tested by simulations for different scenarios. The proposed method assesses whether the experimental data from each participant are consistent with those from the others. Finally, one-dimensional and multidimensional degrees of equivalence (DOE) are defined, which should allow systematic deviations of spectral and geometric nature to be identified.

1. Introduction

The angular dependence of the spectral reflectance must be properly characterized when attempting to evaluate the appearance of objects. Reflectance depends on the directional irradiation and detection vectors ($\mathbf{r}_\mathrm{i}$ and $\mathbf{r}_\mathrm{r}$) and on the projected solid angles of irradiation ($\mathrm{d}\Omega_\mathrm{i}$) and detection ($\mathrm{d}\Omega_\mathrm{r}$), which we will call irradiation/detection geometry hereafter. Its characterization as a quotient between reflected and incident radiant fluxes is only valid in the case of directional Fresnel reflections, where the reflected radiant flux is restricted to the specular direction within a small solid angle of detection. To characterize the reflectance of a surface for any irradiation/detection geometry, the distribution of a bidirectional quantity, which relates the infinitesimal reflected radiant flux from any direction $[\mathrm{d}\Phi_\mathrm{r}(\mathbf{r}_\mathrm{r})]$ and the infinitesimal incident radiant flux from any other direction $[\mathrm{d}\Phi_\mathrm{i}(\mathbf{r}_\mathrm{i})]$, has to be measured. Nicodemus *et al.* proposed the Bidirectional Reflectance Distribution Function (BRDF) in their monograph on reflectance published in 1977 [1]. This has been adopted unanimously by the scientific community to describe the reflectance of surfaces. It is not a quantity as simple as the quotient of radiant fluxes from given directions. The reason is that, given a directional irradiance $E_\mathrm{i}(\mathbf{r}_\mathrm{i})$ on the surface, this mere quotient between fluxes would depend on the projected solid angle of detection, $\mathrm{d}\Omega_\mathrm{r}$, when the reflected flux is not completely uniform within it.

The BRDF was defined [1] as the derivative:

$$f_\mathrm{r}(\mathbf{r}_\mathrm{i};\mathbf{r}_\mathrm{r}) = \frac{\mathrm{d}L_\mathrm{r}(\mathbf{r}_\mathrm{r}, E_\mathrm{i})}{\mathrm{d}E_\mathrm{i}(\mathbf{r}_\mathrm{i})}, \tag{1}$$

which can be expressed as:

$$f_\mathrm{r}(\mathbf{r}_\mathrm{i};\mathbf{r}_\mathrm{r}) = \frac{L_\mathrm{r}(\mathbf{r}_\mathrm{r})}{E_\mathrm{i}(\mathbf{r}_\mathrm{i})} \tag{2}$$

at the domain where the radiance L_r depends linearly on E_i, or on the projected irradiation solid angle $\mathrm{d}\Omega_\mathrm{i}$ ($E_\mathrm{i} = L_\mathrm{i}\mathrm{d}\mathbf{\Omega}_\mathrm{i}$, being L_i the radiance producing irradiation on the surface).

The proper characterization of the spectral BRDF allows reflectance for any irradiation/detection geometry to be calculated by integration [1].

The alternative way to unequivocally define this bidirectional quantity is through the bidirectional radiance factor $[\beta(\mathbf{r}_\mathrm{i};\mathbf{r}_\mathrm{r})]$. The radiance factor (or reflectance factor) is the ratio of the radiant flux actually reflected by a sample surface to that which would be reflected into the same reflected-beam geometry by an ideal (lossless) perfectly diffuser (Lambertian) surface irradiated in exactly the same way as the sample. The bidirectional radiance factor is the radiance factor when the solid angles of irradiation and detection are small enough to fulfill the linearity condition previously defined. There is a proportional relation between $\beta(\mathbf{r}_\mathrm{i};\mathbf{r}_\mathrm{r})$ and $f_\mathrm{r}(\mathbf{r}_\mathrm{i};\mathbf{r}_\mathrm{r})$, by a factor π [1]. In consequence, both bidirectional quantities can be used to characterize the reflectance of surfaces.

In recent years, National Metrology Institutes (NMIs) and other research centers have developed complex robot–based instruments to measure the bidirectional reflectance of surfaces with as few geometrical restrictions as possible [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. Within the activity of the project "Multidimensional reflectometry for industry" funded by the European Metrology Research Program (EMRP), an interlaboratory comparison is in process to assess the calibration and measurement capabilities (CMC) of some of the participating centers to measure the spectral BRDF or, equivalently, the spectral bidirectional radiance factor. Given the complex nature of spectral BRDF data, which depends on the spectral variable and on the geometric variables corresponding with irradiation and detection directions, a non-standard comparison analysis procedure is required. It is not sufficient to consider comparison analysis based on a single measurand, nor to use a methodology based on determining measurement curves and comparison reference curves, as proposed in Ref. [12], since such a procedure is applied for data smoothly depending on a single variable.

A straightforward data analysis method for the comparison of data produced by the different participants was established within the above–mentioned project and it is described in this article. It assesses whether the experimental data from each participant are consistent with the experimental data from the other participants. We test the method with simulations at some scenarios to understand its sensitivity and ability to identify inconsistent data. Finally, the method defines one-dimensional and multidimensional degrees of equivalence (DOE), which should allow systematic deviations to be identified.

2. Methodology

The aim of this comparison is to establish the consistency of the multidimensional measurements provided by the participants, but also to identify systematic spectral or geometric errors. In order to test the consistency, we will use statistical methods to accept or reject a hypothesis of non difference among groups, which is usually called null hypothesis (H_0) when the hypothesis is that chance alone is responsible for the results. On the other hand, to identify systematic errors, we will calculate one-dimensional and multidimensional degrees of equivalence (DOEs) (deviations of the measured values provided by the participants from a reference value). One-dimensional DOEs are usually provided for comparisons with a small number of measurands. We propose here the definition of multidimensional DOEs to give insight into the structure of data in both spectral and geometric dimensions.

To test the consistency of a participant, the null hypothesis H_0 is that there is not a statistical difference between the measurement provided by that participant and a comparison reference value (CRV) representing the measurements of all participants. H_0 can be formally expressed as:

$$\mathbf{M}_0 = \mathbf{M}_\mathrm{k} \tag{3}$$

where $\mathbf{M}_\mathrm{k}$ is the array containing the experimental data of the k-participant, and $\mathbf{M}_0$ is the array containing the CRV (comparison reference value) for each measurement configuration. Since we have N null hypotheses, we may use multiple hypothesis testing (e.g., the Bonferroni procedure and the procedures described in Refs. [13, 14] to evaluate if one of these hypotheses was rejected by chance, which can happen with a large number of participants.

A 95% confidence interval is used to reject or not the hypothesis. Thus, the significance level (probability of obtaining a larger statistical difference) is p = 0.05. In other words, if the obtained p-value is lower than 0.05, the deviation from the null hypothesis is statistically significant and the hypothesis is rejected.

The statistical difference between $\mathbf{M}_0$ and $\mathbf{M}_\mathrm{k}$ is accounted by the chi-squared random variable:

$$\chi^2(k) = \sum_{\mathrm{i}=1}^{d} \frac{(\tilde{M}_\mathrm{i} - M_{\mathrm{k,i}})^2}{\sigma_\mathrm{i}^2} \tag{4}$$

where $\tilde{\mathbf{M}}$ is the estimation of $\mathbf{M}_0$, σ_i^2 represents the variance of $\tilde{M}_\mathrm{i} - M_{\mathrm{k,i}}$, the i-subscripts represent every configuration (fixed wavelength and irradiation/detection geometry), and d is the number of measurement configurations. The value of σ_i is the quadratic sum of an estimated uncertainty for $\tilde{\mathbf{M}}$ $[\sigma(\mathbf{M}_0)]$ and the standard uncertainty provided by the participants. Because of the data complexity, with both spectral and geometric dependence, it is complicated to consider, as usually, covariance matrix of data to account for correlation between quantities [15].

2.1. Estimation of the CRV

Although the weighted mean is recommended in high level comparisons [16], the functional form of the CRV ($\mathbf{M}_0$) is not specified by the mutual recognition arrangement (CIPM MRA) in 1999 [17] and its definition is still controversial, as discussed in [18]. Since we have to deal with a low number of participants in the comparison and no previous testing of the uncertainty budgets has been done, we decided to estimate $\mathbf{M}_0$ simply as the unweighted mean of the experimental data of all participants:

$$\tilde{\mathbf{M}} = \langle \mathbf{M}_\mathrm{k} \rangle = \frac{1}{N} \sum_{\mathrm{k}=1}^{N} M_\mathrm{k} \tag{5}$$

where N is the number of participants.

$\sigma(\mathbf{M}_0)$ represents the uncertainty of this estimation, and it is calculated from the data variance divided by the number of participants (variance of the mean), as:

$$\sigma(\mathbf{M}_0) = \sqrt{\frac{\left\langle {\mathbf{M}_\mathrm{k}}^2 \right\rangle - \langle \mathbf{M}_\mathrm{k} \rangle^2}{N}} \tag{6}$$

2.2. Assessment of consistency for a given participant

The chi-squared expression shown in Eq. 4 is a quadratic sum of independent standard normal random variables. The expected value of the numerators of the terms in this equation coincides with the variance and, therefore, the expected value of χ^2 coincides with the number of dimensions of the complete measurement (d). Therefore, given a calculated value of $\chi^2(k)$ for the measurement of the k-participant, the probability that χ^2 was higher than $\chi^2(k)$ is calculated from the chi-squared distribution. This probability is the p-value. If it is less than the significance level 0.05, the deviation from the null hypothesis is statistically significant and the measurement of the k-participant is considered inconsistent.

The measurement of the spectral bidirectional radiance factor or the spectral BRDF at d_g irradiation/detection geometries and at d_λ wavelengths is a set of $d_\mathrm{g} \times d_\lambda$ measures (a d-dimensional measure $\mathbf{M}$ with $d = d_\mathrm{g} \times d_\lambda$). We define the random variable that accounts for the statistical difference between $\mathbf{M}_0$ and $\mathbf{M}_\mathrm{k}$ at a fixed wavelength λ as:

$$\chi^2_\lambda(k) = \sum_{\mathrm{i}=1}^{d_\mathrm{g}} \frac{(\tilde{M}_\mathrm{i} - M_{\mathrm{k,i}})^2}{\sigma_\mathrm{i}^2} \tag{7}$$

where the summation is done just for the data with wavelength λ. In this case, we would obtain d_λ p-values, which can be represented as a spectral plot.

Similarly, the statistical difference between $\mathbf{M}_0$ and $\mathbf{M}_\mathrm{k}$ at a fixed geometry g is:

$$\chi^2_\mathrm{g}(k) = \sum_{\mathrm{i}=1}^{d_\lambda} \frac{(\tilde{M}_\mathrm{i} - M_{\mathrm{k,i}})^2}{\sigma_\mathrm{i}^2} \tag{8}$$

where the summation is done just for the data with a given geometry g. In this case, we would obtain d_g p-values, which can be represented in some kind of plot used to represent geometry-related variations.

These definitions of χ^2_g and χ^2_λ are useful to identify the source of systematic deviations. For instance, some conclusion can be drawn from the spectral representation of the p-values for the different wavelengths (χ^2_λ-based analysis). In short, if the calculated p-value is independent of the wavelength, any systematic spectral deviation is negligible with respect to other errors sources. A similar conclusion can be drawn from the χ^2_g-based analysis.

The final comparison between laboratories may provide a single value p-value of the complete measurement, and two sets of d_g and d_λ p-values at fixed geometric and spectral conditions.

2.3. Assessment of DOE and systematic deviations

The simplest way to assess systematic deviations in measurements is by studying the variations of the relative difference between $\tilde{\mathbf{M}}$ and $\mathbf{M}_\mathrm{k}$. The values calculated as:

$$\Delta_\mathrm{i}(k) = \frac{\tilde{M}_\mathrm{i}}{M_{\mathrm{k,i}}} - 1 \tag{9}$$

represent these differences for every i-configuration and it is a d-dimensional DOE.

Now, if the geometrically–averaged values of Δ_{i} are calculated at fixed wavelengths as:

$$\langle \Delta_\lambda \rangle (k) = \frac{1}{d_{\mathrm{g}}} \sum_{\mathrm{i}=1}^{d_{\mathrm{g}}} \Delta_{\mathrm{i}}(k) \tag{10}$$

the systematic deviation with spectral dependence can be quantified. And, similarly, to quantify the geometric dependence of a possible systematic deviation:

$$\langle \Delta_{\mathrm{g}} \rangle (k) = \frac{1}{d_\lambda} \sum_{\mathrm{i}=1}^{d_\lambda} \Delta_{\mathrm{i}}(k) \tag{11}$$

The standard uncertainty associated with the deviation $\Delta_{\mathrm{i}}(k)$ is calculated as:

$$u(\Delta_{\mathrm{i}})(k) = \sqrt{\sigma^2(\mathbf{M}_0) + u^2_{\mathrm{rep,k,i}}} \tag{12}$$

where $u^2_{\mathrm{rep,k,i}}$ is the reported standard uncertainty by the k-participant at the i-configuration. Form these uncertainties, the standard uncertainties associated with the deviations $\langle \Delta_\lambda \rangle (k)$ and $\langle \Delta_{\mathrm{g}} \rangle (k)$ are calculated as:

$$u(\Delta_\lambda)(k) = \sqrt{\frac{\sum_{\mathrm{i}=1}^{d_{\mathrm{g}}} u^2(\Delta_{\mathrm{i}})(k)}{d_{\mathrm{g}}}} \tag{13}$$

and

$$u(\Delta_{\mathrm{g}})(k) = \sqrt{\frac{\sum_{\mathrm{i}=1}^{d_\lambda} u^2(\Delta_{\mathrm{i}})(k)}{d_\lambda}}, \tag{14}$$

respectively.

We will show the performance of this quantification in the next section.

3. Numerical evaluation

To evaluate the performance of the proposed method, we have simulated spectral BRDF measurements for four artificial participants (Part_1, Part_2, Part_3 and Part_4). We have given these participants different systematic errors to understand the sensitivity of the approach to these types of errors. The experimental BRDF of a special effect coating with Merck Colorstream Viola Fantasy pigments (angle–dependent color travel from lilac through silver and green to blue), as measured by the goniospectrophotometer GEFE at IO-CSIC [7], is considered as the true value (reference value) in this simulation. To show the high diversity of the spectral BRDF of this coating, the data at four selected geometries are plotted in Fig. 1. θ_{i} and θ_{r} are the polar angles of irradiation and detection, respectively, whereas ϕ_{i} and ϕ_{r} correspond with azimuth angles. We defined that always $\phi_{\mathrm{i}} = 0^{\mathrm{o}}$, and, therefore, the value of ϕ_{r} represents the difference between the azimuth angles of the irradiation and detection directions. Aspecular angles is the angular deviation between the detection direction and the specular direction. We used eight polar angles for both irradiation and detection (from θ_{i} and $\theta_{\mathrm{r}} = 0^{\mathrm{o}}$ to 70^{o}, with angular steps of 10^{o}), and only two azimuth angles for detection ($\phi_{\mathrm{r}} = 0^{\mathrm{o}}$ and 180^{o}),

which define incidence plane. Thus, we obtained $d_g = 128$ geometries. On the other hand, the wavelength variable took values from 380 nm to 780 nm, with steps of 1 nm ($d_\lambda = 401$ nm).

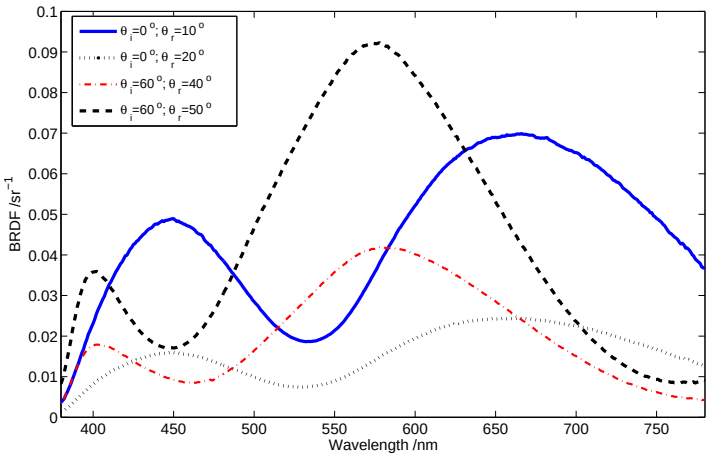

Figure 1. Experimental spectral BRDF used as true value for the simulations (special effect coating with Merck Colorstream Viola Fantasy pigments). The data at four selected geometries (one high irradiation angle [60°], one low irradiation angle [0°], and two low aspecular geometries [10°] and [20°]) are plotted versus wavelength.

For the sake of simplicity we will assume that all participants reported a standard uncertainty of 0.4 %. In addition to this uncertainty (treated as Gaussian noise) we have introduced three types of systematic errors that are not accounted for in their uncertainty budgets. These, multiplied by the reference value, will produce the erroneous measurements. These are:

(i) Constant systematic error (C_{sys}) of 0.9955.

(ii) Geometric systematic error (G_{sys}) as shown in Fig. 2 for some geometries, whose value increases proportionally to the root of the norm of the difference vector between the direction vectors at a given geometry and the 0° : 45° geometry.

(iii) Spectral systematic error (S_{sys}) as shown in Fig. 3, whose value increases proportionally to the root of the absolute difference between the given wavelength and 580 nm.

From here, four independent scenarios were devised with different systematic errors for every participant, as detailed in Table 1. Random errors were also introduced, according to the uncertainty budget. In the first three scenarios one systematic error type was attributed to participants 2, 3 and 4 in turn, the others being free of error. In the fourth scenario, participants 2, 3 and 4 all had systematic errors (one of each kind) and participant 1 was free of error.

The result of the consistency test for every scenario are given in Table 2 for all participants (inconsistent if $p<0.05$). As expected, if a participant is free from

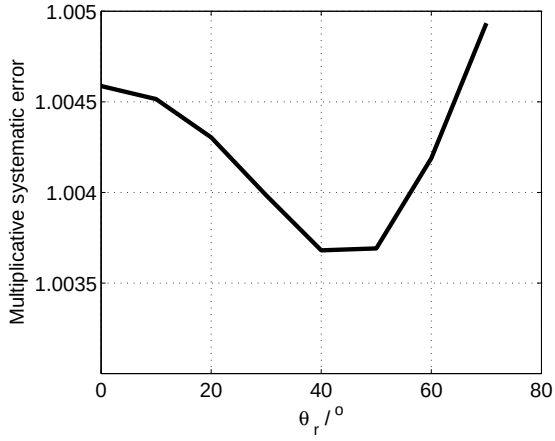

Figure 2. Multiplicative geometric systematic error for geometries at $\theta_{\mathrm{i}} = 0^{\mathrm{o}}$ and $\phi_{\mathrm{r}} = 180^{\mathrm{o}}$.

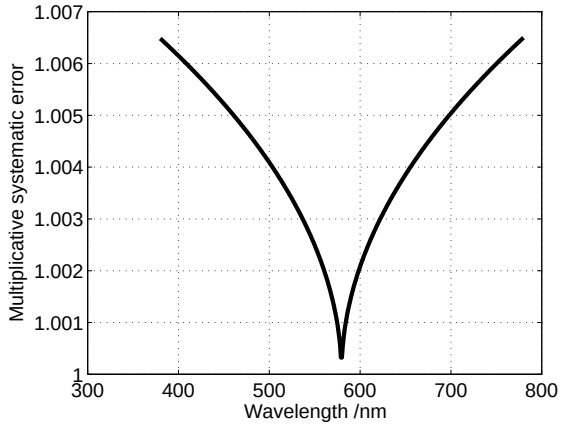

Figure 3. Multiplicative spectral systematic error.

Table 1. Systematic errors used in the different scenarios.

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Part_1	-	-	-	-
Part_2	-	-	$\mathrm{S}_{\mathrm{sys}}$	$\mathrm{S}_{\mathrm{sys}}$
Part_3	-	$\mathrm{G}_{\mathrm{sys}}$	-	$\mathrm{G}_{\mathrm{sys}}$
Part_4	$\mathrm{C}_{\mathrm{sys}}$	-	-	$\mathrm{C}_{\mathrm{sys}}$

systematic errors, its measurement will be consistent with the measurements of the other participants. However, it is not always the case that a participant with a systematic error has an inconsistent output. In the fourth experiment, when three out of four participants have some kind of systematic error and, in consequence, the uncertainty $\sigma(\mathbf{M}_0)$ is high (Eq. 6), only one participant is labeled as inconsistent.

The p-values at fixed geometries obtained from the evaluation of $\chi^2_{\mathrm{g}}(k)$ (Eq. 8) and

Table 2. Results of the consistency test for the different scenarios (inconsistent if p<0.05).

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Part_1	consistent	consistent	consistent	consistent
Part_2	consistent	consistent	**inconsistent**	consistent
Part_3	consistent	**inconsistent**	consistent	consistent
Part_4	**inconsistent**	consistent	consistent	**inconsistent**

at fixed wavelengths (from $\chi^2_\lambda(k)$, in Eq. 7) are shown in Figs. 4 (versus the detection angle, for geometries at $\theta_\text{i} = 0^\text{o}$ and $\phi_\text{r} = 180^\text{o}$) and 5 (versus wavelength), respectively.

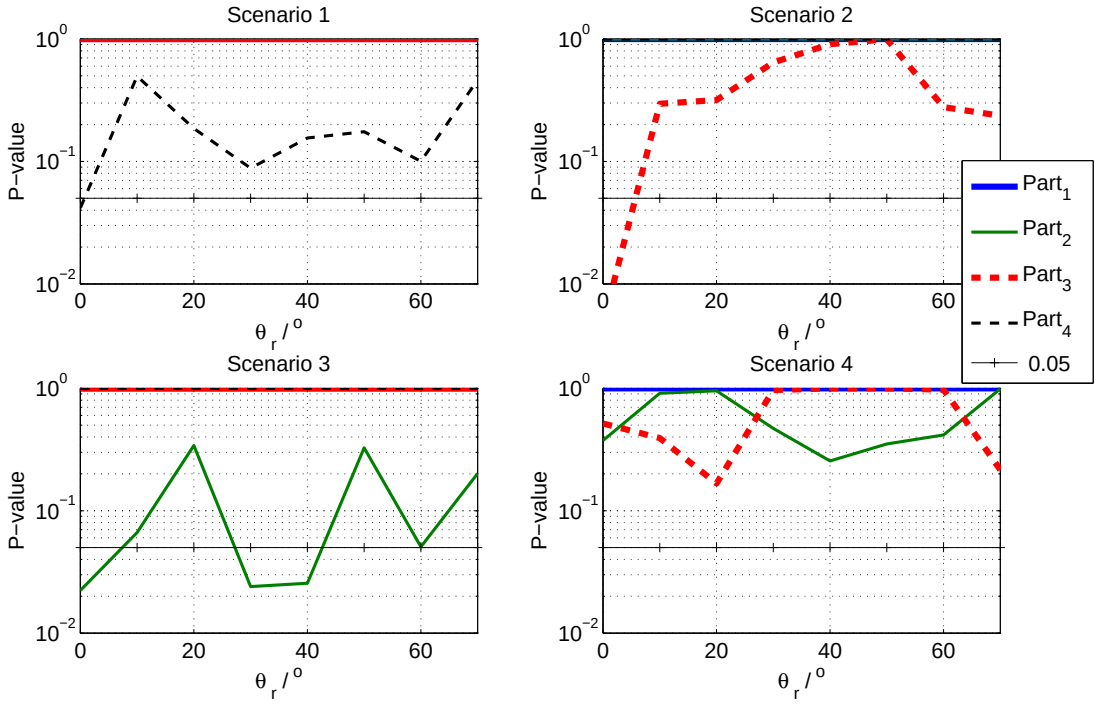

Figure 4. Participants' p-values at fixed geometries (Eq. 8) versus the detection angle, for every scenario studied in this work (only shown those at $\theta_\text{i} = 0^\text{o}$ and $\phi_\text{r} = 180^\text{o}$). Curves with p-values of 1 are not shown in the plots, nor that the corresponding to the fourth participant in scenario 4, whose p-value is lower than 0.01.

Except for some geometries, all the p-values obtained from the χ^2_g-based analysis (in Fig. 4) lie above the line corresponding to the p-value = 0.05 for all scenarios, which means that the results are consistent. It must be noticed that only those participants with systematic errors in scenarios 1, 2 and 3 have p-values appreciably below 1. In scenario 2, for which Part_3 has a geometric systematic error, it can be observed that the consistency of this participant is the higher (p-value closer to 1) the nearer the geometry

is to the $0^o : 45^o$ geometry, revealing the nature of the introduced systematic error (see Fig. 2). In scenario 4, for which three laboratories have different systematic errors, the consistency of Part_2 and Part_3 improves with respect to scenarios 2 and 3, due to the higher uncertainty in the estimation of $\mathbf{M}_0$ (Eq. 6). However, the measurement of Part_4 turns completely inconsistent (p-value < 0.01, and not shown in the figure). The reason is that the systematic error introduced for this participant reduces its value, unlike the other systematic errors in Part_2 and Part_3, which increase them. So the estimation of $\mathbf{M}_0$ increases respect to the true value, in contrast to the value provided by the fourth participant.

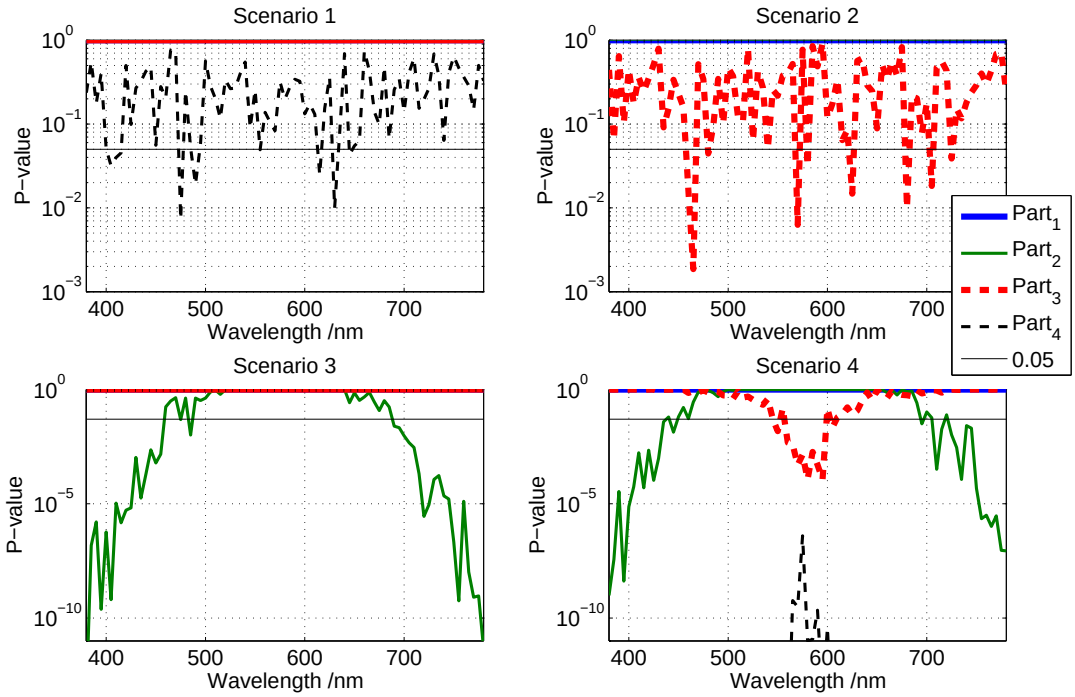

Figure 5. Participants' p-values at fixed wavelengths (Eq. 7) versus wavelength, for every scenario studied in this work. Curves with p-values of 1 are not shown in the plots.

p-values obtained from the χ^2_λ-based analysis are shown in Fig. 5. Scenarios 1 and 2 show noisy spectral structures for the participants whose results are affected by constant (Part_4 in scenario 1) and geometric (Part_3 in scenario 2) systematic errors. This structure is due to the geometrical behavior of the spectral BRDF. Then, generally, a noisy feature for the p-value is related to a systematic error independent of the wavelength. However, in scenario 3, it can be observed that the consistency of Part_2 is well-correlated with the spectral systematic error introduced (see Fig. 3). Scenario 4 shows spectral features for the three participants affected by systematic errors, although only one of them has a spectral systematic error. This is because of the change in the reference value, similarly as it will be shown in Figs. 6 and 7.

The consistency analysis based on χ^2_g and χ^2_λ give some clues about the presence

of systematic errors in participants, but, to better quantify them, the multidimensional DOE (Eqs. 11 and 10) can be used. They are represented in Figs. 6 and 7 for the fourth scenario, which is the most problematic under the consistency analysis.

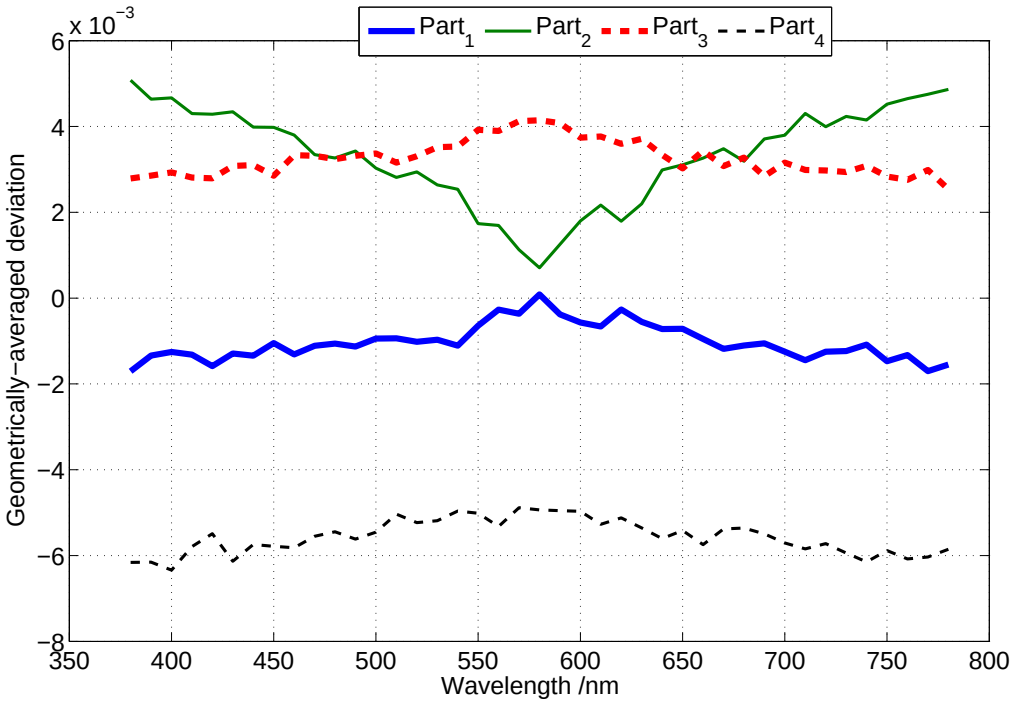

Figure 6. Geometrically-averaged deviation at fixed wavelength of every participant, versus wavelength, for the 4th scenario described in the text.

In Fig. 6, it is clearly observed the spectral systematic error of Part_2 producing an opposite systematic effect in the other participants, but around three times smaller, because of the averaging over the number of participants. Notice that, whereas the maximum variation across the wavelength range of the geometrically-averaged deviation is between 10^{-3} to 1.5×10^{-3} for Part_1, Part_3 and Part_4, it is around 5×10^{-3} for Part_2. This differences are larger than the standard uncertainties associated with the deviations $\langle \Delta_\lambda \rangle (k)$ (Eq. 13), which are around 4×10^{-4}.

In Fig. 7, which shows the DOE as a function of angle for some geometries for scenario 4, it is observed that all participants have a deviation slightly dependent on the detection angle, except for Part_3 (with geometric systematic error), whose deviation is the lower the nearer the geometry is to the $0^o : 45^o$ geometry, where the systematic error is null by definition. Again, a differential feature in the graph points out a systematic error of one participant.

4. Conclusion

A data analysis method for the comparison of spectral BRDF data produced by different participants has been presented. This method deals with a low number

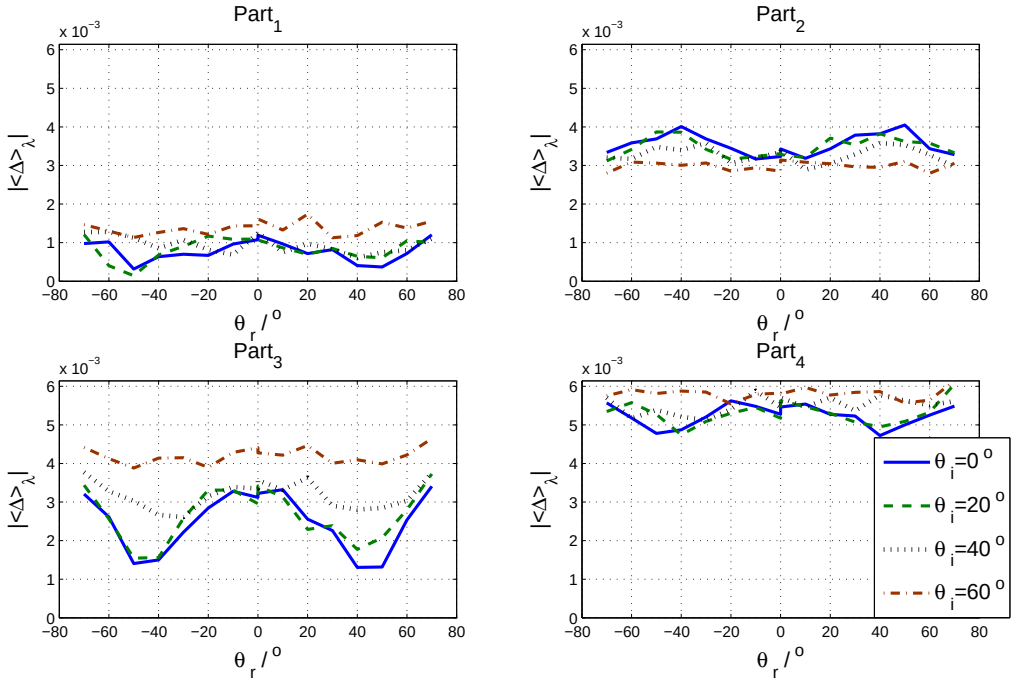

Figure 7. Spectrally-averaged deviation at fixed geometry of every participant, versus detection angle, for the 4th scenario described in the text .

of participants and with multidimensional measurements. It assesses whether the experimental data from each participant are consistent with the experimental data from the other participants. One-dimensional and multidimensional degree of equivalence (DOE) quantities have been defined. To evaluate the performance of the proposed methods, we have simulated spectral BRDF measurements for four participants using constant, geometric and spectral multiplicative systematic errors. By using these simulated BRDF measurements, it has been shown that the method allows the presence and type of systematic deviations to be identified in spectral or geometric plots.

Acknowledgements

This report was compiled within the EMRP IND52 Project xD-Reflect "Multidimensional reflectometry for industry". The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union. Part of the authors (Instituto de Óptica "Daza de Valdés" (IO-CSIC), Agencia Estatal CSIC) are also grateful to Comunidad de Madrid for funding the project SINFOTON-CM: S2013/MIT-2790.

References

[1] Nicodemus F E, Richmond J C and Hsia J J 1977 *Geometrical considerations and nomenclature for reflectance* (Natl. Bur. Stand. Monogr. **160**)

[2] Germer T A and Asmail C C 1999 *Rev. Sci. Instrum.* **70** 3688–3695

[3] Nevas S, Manoocheri F and Ikonen E 2004 *Appl. Opt.* **43** 6391–6399
[4] Hünerhoff D, Grusemann U and Höpe A 2006 *Metrologia* **43** S11–S16
[5] Leloup F B, Forment S, Dutré P, Pointer M R and Hanselaer P 2008 *Appl. Opt.* **47** 5454–5467
[6] Baribeau R, Neil W S and Côté E 2009 *Journal of Modern Optics* **56** 1497–1503
[7] Rabal A M, Ferrero A, Campos J, Fontecha J L, Pons A, Rubiño A M and Corróns A 2012 *Metrologia* **49** 213–223
[8] Matsapey N, Faucheu J, Flury M and Delafosse D 2013 *Meas. Sci.Technol.* **24** 065901
[9] Ouarets S, Ged G, Razet A and Obein G 2012 A new gonioreflectometer for the measurement of the bidirectional reflectance distribution function (brdf) at LNE-CNAM (*Proceedings of CIE 2012 "Lighting Quality and energy efficiency"* vol 5880) pp 687–691
[10] Obein G, Bousquet R and Nadal M E 2005 New NIST reference goniospectrometer (*Proc. SPIE* vol 5880) pp 241–250
[11] Patrick H J, Zarobila C J and Germer T A 2013 The NIST robotic optical scatter instrument (ROSI) and its application to brdf measurements of diffuse reflectance standards for remote sensing SPIE Optical Engineering+Applications. International Society for Optics and Photonics pp 886615–886615–12
[12] Cox M G and Harris P M 2012 *Metrologia* **49** 437–445
[13] Yoav Benjamini Y H 1995 *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 289–300 ISSN 00359246 URL `http://www.jstor.org/stable/2346101`
[14] Benjamini Y and Yekutieli D 2001 *Ann. Statist.* **29** 1165–1188 URL `http://dx.doi.org/10.1214/aos/1013699998`
[15] Nielsen L 2003 *Measurement Techniques* **46** 513–522 ISSN 0543-1972
[16] CIPM MRA-D-05 2014 Measurement comparisons in the CIPM MRA, version 1
[17] Bureau International des Poids et Mesures 1999 Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes
[18] Elster C and Toman B 2013 *Metrologia* **50** 549–555